Comparison of the microcanonical population annealing algorithm with the Wang-Landau algorithm

Vyacheslav Mozolenko[1,2], Marina Fadeeva[2], and Lev Shchur[1,2]

[1] *Landau Institute for Theoretical Physics, 142432 Chernogolovka, Russia and*

[2] *HSE University, 101000 Moscow, Russia*

The development of new algorithms for simulations in physics is as important as the development of new analytical methods. In this paper we present a comparison of the recently developed microcanonical population annealing (MCPA) algorithm with the rather mature Wang-Landau algorithm. The comparison is performed on two cases of the Potts model exhibiting a first order phase transition. We compare the simulation results of both methods with exactly known results, including the finite-dimensional dependence of the maximum of the specific heat capacity. We evaluate the Binder cumulant minimum, the ratio of peaks in the energy distribution at the critical temperature, and the energies of the ordered and disordered phases. Both methods exhibit similar accuracy at selected sets of modeling parameters.

INTRODUCTION

Numerical methods for investigating statistical mechanical models consist of a wide range approaches, among them [47] low- and high-temperature series [1, 2], transfer-matrix methods [3, 4], Markov Chain Monte Carlo (MCMC) [5, 6], cluster Swendsen-Wang and Wolff methods [7, 8], multicanonical methods [9, 10], Wang-Landau method [11, 12], simulated annealing methods [13], population annealing methods [14, 15], etc. A recent systematic review of methods can be found in Ref. [16]. A common feature of these approaches is the explicit temperature dependence of the modeling process, with the exception of the Wang-Landau method. Temperature dependence leads to a critical slowing down of the simulation in the region around the critical point of the continuous transition [5], which leads to a power law dependence of the correlation time on the system size, $t \propto L^z$, with the value of z typically around 2 for local MCMC [17] and an order of magnitude lower for cluster algorithms [18]. The practical computation time of a d-dimensional system, measured as the number of operations per spin, grows as L^{d+z}. Modeling of discontinuous phase transitions in the critical region is determined by the interface between the regions of ordered and disordered phases. The surface energy of which is proportional to the length of the interface, forming an energy barrier that depends exponentially on the length of the interface. This makes the simulation unpredictably slow in the critical region.

The Wang-Landau algorithm has no temperature-dependent probabilities and uses the current estimate of the density of states (DoS) to calculate the probability of transition from one energy level to another. This algorithm relaxes system to the true DoS [19]. The relaxation process converges steadily with an appropriate modification [20] of the original Wang-Landau algorithm and convergences as slowly as $1/t$ [21], with the time t measured in elementary steps of a single-spin flip.

Another algorithm that has no temperature-dependent probabilities [23] was recently presented, the microcanonical population algorithm. The special feature of this algorithm, is that it does not use any probabilities in the annealing part. Instead, it uses the *ceiling* of energy when moving down the energy spectrum or the *floor* of energy when moving up the energy spectrum [24]. The next feature is the parallel simulation of a huge number of replicas at a given energy, which allows us to estimate the energy dependence of thermodynamic functions. Together with the DoS estimation, this gives their temperature dependence.

It seems that there is no critical slowdown in the usual sense in microcanonical algorithms due to the absence of an explicit temperature dependence of the algorithm on the temperature. In the case of the first-order phase transition, the condensation and evaporation of droplets in the vicinity of the critical temperature are still determined by the energy barrier associated with surface tension and depend exponentially on the surface length. It is likely that microcanonical simulations is less sensitive to this than the canonical simulations [23, 25]. There are example of the comparison of energy barriers in the Lennard-Jones particle system showing the same scaling behaviour but barrier height is almost six times smaller for the microcanonical ensemble comparing with the canonical one [26].

In this paper, a comparative study of two algorithms, modified Wang-Landau algorithm (WL-1/t) [11, 20] and the multicanonical population algorithm (MCPA) [23, 24], is carried out. The reason for the comparative study is to understand accuracy of the methods. As an example, we simulate Potts model with 10 and 20 components that undergoes a strong discontinuous transition. We use the elementary step of a single-spin flip as the Monte Carlo step. The elementary step is very different in the two models due to the intrinsic difference of the models. We model one replica of the system with the WL-1/t algorithm, and R parallel replicas of the system with the

MCPA algorithm. So, such defined Monte Carlo steps are more or less proportional to the simulation time. The details are discussed in the conclusion section of the paper.

MODEL

The Potts model [27, 28] with q-state is a generalization of the Ising model consisting of N interacting spins σ_i ($i=1,...,N$), each of which taking values $\sigma_i \in (1,...,q)$. The energy E is given by

$$E = -J \sum_{\langle i,j \rangle} \delta(\sigma_i, \sigma_j) \tag{1}$$

and for simplicity we set the ferromagnetic constant J equal to unity $J=1$.

The critical temperature is known analytically (see [29])

$$\beta_c = \frac{J}{k_B T_c} = \ln(1 + \sqrt{q}). \tag{2}$$

For $q > 4$, the Potts model undergoes a first order phase transition and a mixing of ordered and disordered phases is observed between the respective energies e_o and e_d. In the thermodynamic limit $L \to \infty$ these energies are also known (see [29])

$$\frac{e_o + e_d}{2} = -\left(1 + \frac{1}{\sqrt{q}}\right) \tag{3}$$

and

$$e_d - e_o = 2\left(1 + \frac{1}{\sqrt{q}}\right) \tanh\left(\frac{\Theta}{2}\right) \prod_{n=1}^{\infty} \tanh^2(n\Theta), \tag{4}$$

where Θ is defined as $2\cosh(\Theta) = \sqrt{q}$.

DIRECT ESTIMATION OF DOS

Procedures of calculating thermodynamic quantities from the density of states (DoS) are based on the representation of a partition function, replacing the summation over all possible spin configurations $\{j\}$

$$Z = \sum_{\{j\}} e^{-E(\{j\})/k_B T}, \tag{5}$$

where k_B is the Boltzmann constant, T is the temperature, with the summation over the energy levels n

$$Z = \sum_{\{n\}} g(E_n) e^{-E_n/k_B T}. \tag{6}$$

The function $g(E)$ is the density of states (DoS), that is, $g(E_n)$ is the number of configurations with energy E_n.

The are two algorithms for direct estimation of DoS (technically, the logarithm of DoS), the Wang-Landau algorithm [11, 12] and the microcanonical population annealing algorithm [23, 24], which are the subject of a comparative study in this paper.

Wang-Landau algorithm

The Wang-Landau algorithm [11, 12] is applicable to any system with a given partition function, which can be written as a sum over n energy levels, as in Expr. (6). This representation is key to the Wang-Landau algorithm. Knowing $g(E)$, one may calculate, for example, the internal energy $E(\beta)$ and heat capacity $C(\beta)$ at any value of inverse temperature $\beta = 1/k_B T$.

$$E(\beta) = \langle E \rangle = \frac{\sum_{n=0}^{N_E - 1} E_n g(E_n) e^{-\beta E_n}}{\sum_{n=0}^{N_E} g(E_n) e^{-\beta E_n}}, \tag{7}$$

$$C(\beta) = \beta^2(\langle E^2 \rangle - \langle E \rangle^2). \tag{8}$$

The basic idea of the algorithm is to organize a random walk through the energy space of the system using an appropriate transition probability for a random walk jump from level energy k to energy level m, this Wang-Landau probability defined below in expr. (9).

During the random walk [11], two histograms $H(E)$ and $S(E)$ are accumulated. The first histogram $S(E) = \log(g(E))$ contains the current value of the DoS logarithm. The second is the auxiliary histogram $H(E)$, which contains information about the number of visits to each energy level. At the beginning of the algorithm, $H(E_n)$ is initialized with zeros and $S(E_n)$ with ones. The initial configuration of the system is set to the ground state. The initial value of the modification parameter $f = f_0 = \exp(1) \simeq 2.71828$.

The further steps of the algorithm are as follows: 1) the state of the system is changed (for the spin model it is a flip of the random spin) and the energy of the new state E_m is calculated; 2) the transition from the state with energy E_k to the state with energy E_m takes place with the Wang-Landau probability

$$P_{WL}(E_k, E_m) = \min\left(1, \frac{\tilde{S}(E_k)}{\tilde{S}(E_m)}\right), \tag{9}$$

where $\tilde{S}(E_k) = \log(\tilde{g}(E_k))$ is the logarithm of the current DoS estimate. The next step is to update the auxiliary histogram $H(E_k) \to H(E_k)+1$, and the current estimate $\tilde{S}(E_k) \to \tilde{S}(E_k) + \ln(f)$.

Steps 1) and 2) are repeated until the histogram $H(E_k)$ is sufficiently "flat", e.g., at the 5% level [11]). The value of the parameter f is then updated as a function of the square root $f_i = \sqrt{f_{i-1}}$, and the histogram is reset $H(E_k)=0$. Steps 1) and 2) are then repeated again. The

algorithm terminates when the parameter f reaches some desired value of f_{end}, e.g., $f_{fin} = \exp(10^{-8}) \simeq 1.00000001$, as proposed in [11].

1/t-Wang-Landau algorithm

The Wang-Landau algorithm is used in various fields of science, for example, in modeling polymers [30, 31] and protein chains [32, 33], for optimization [21]. However, it is known that as the number of steps increases, the computational accuracy saturates at some step. This has been observed in a number of publications such as [34, 35]. Consequently, it affects the accuracy of the calculation of thermodynamic functions for relatively large systems, including in the critical region. In addition, the choice of $\sqrt{f_i}$ as the function to change the parameter f was left open until proposed in [20] modification.

The 1/t-Wang-Landau algorithm is a modification [20] that emerged as a solution to these problems. The first step of 1/t-Wang-Landau is similar to the original [11] method, except that criterion of histogram $H(E)$ "flatness" was replaced by the $H(E) \neq 0$ test. It is argued in [20] that practically the flatness criteria can be replaced by a weaker but still global property, which is the requirement to visit each energy level before update the parameter f.

The algorithm proceeds to the second stage when the condition $F_i > N_E/t$ is satisfied, where $F = \ln f$. From this point on, the parameter F changes not as $F_i = F_{i-1}/2$ but $F_i = N_E/t$. Here, t is the number of elementary spin flips and N_E is the number of levels in the system. At the second stage the histogram $H(E)$ is no longer checked. Theoretical justification of the convergence of this method was presented in the article [21].

Criterion of convergence

In the original Wang-Landau algorithm [11] and its modification 1/t-Wang-Landau algorithm [20] no method was proposed to evaluate the accuracy and convergence of the algorithm. The only argument for the applicability of the method is to demonstrate the quality of the data by comparing them with the exact solution known for the 2d Ising model [36]. A possible solution to this problem was proposed in [19], where a square matrix $T(E_k, E_m)$ was additionally introduced instead of the histogram $H(E)$, and it was shown that the closeness of the matrix T to stochastic can be used as an accuracy criterion for DoS estimation. The matrix elements are the transition frequency from configurations with energy E_k to configurations with energy E_m. The closer the estimated DoS is to the true DoS, the closer the matrix T is to the stochastic one, and the deviation of the largest eigenvalue of the

matrix from unity can be used as a criterion for the accuracy of the DoS estimate. Interestingly, this property of the matrix T is consistent with the criteria of flatness of the histogram $H(E)$ proposed by Wang and Landau and with the criterion of non-zero visitation of each energy level proposed by Belardinelli and Pereyra. Thus, the matrix T is more informative and, in addition, provides a clue for evaluating the accuracy and controlling the convergence of DoS.
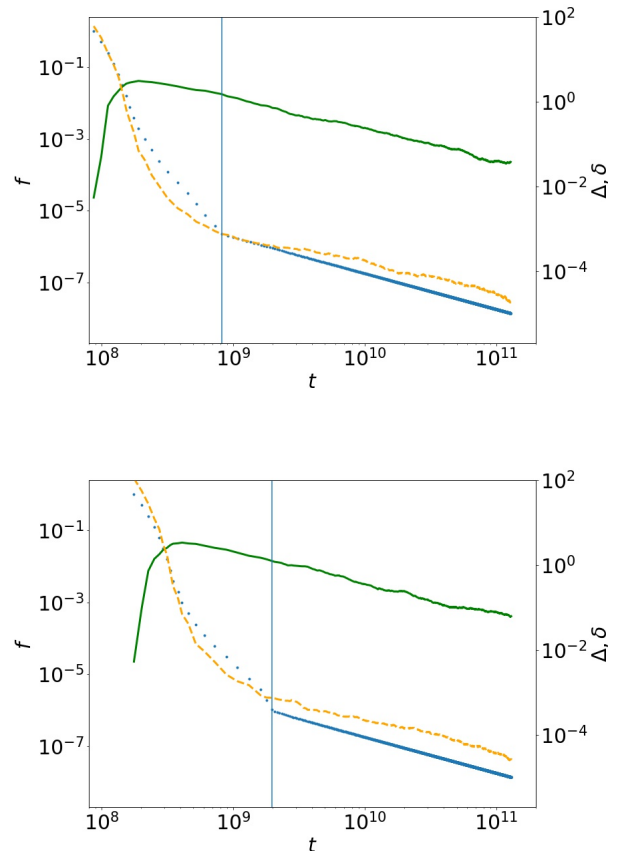

FIG. 1: Dependence of the algorithm parameter f (dotted blue line and left axis), convergence criteria δ (solid green line and right axis) and $\tilde{\Delta}$ (dashed yellow line and right axis) on the Monte-Carlo step t for 10 state Potts model (top panel) and 20 state Potts model (bottom panel). Linear lattice size $L = 30$. The values of f, δ and $\tilde{\Delta}$ are the average over 20 runs. The vertical line is the average time t_0 for which the algorithm switches the function of the modification parameter f.

The figures 1 show the variation of the algorithm parameter f, which is varied according to the Wang-Landau [11] rule $f_i = \sqrt{f_{i-1}}$ until the Belardinelli and Pereyra [20] criteria is met and then by the $f = f_0/t$ rule. Also shown in the figures 1 is the difference Δ, which is calculated as the relative difference of DoS at simulation

time t, $g(E_n;t)$ and at the final time t_f, $g(E_n;t_f)$

$$\Delta = \frac{1}{N_E} \sum_{n=1}^{N_E} \left(\frac{g(E_n;t)}{g(E_n;t_f)} - 1 \right) \qquad (10)$$

and the convergence criterion δ [19]

$$\delta = |1 - \lambda_1|, \qquad (11)$$

where λ_1 is the largest eigenvalue of the transition matrix T, and the two simulation sets presented in Figure 1 refer to the 10 and 20 components of the Potts model.

The parameter Δ is well defined for models for which we have an exact solution for DoS, such as the Ising model in one and two dimensions. In our case with unavailable exact DoS, we replace the unknown exact DoS values with the final DoS values $g(E_n;t_f)$ from simulations. Time is measured in units of single-spin flip events. It can be seen that Δ decreases inversely with time, as shown by Liang [21], and the control parameter δ decreases in the same way [19]. It should be noted that the δ convergence criterion can be used without knowledge of the DoS function and at any time step in the simulation.

The $1/t$ dependence appears to be the optimal protocol for changing the algorithm parameter f [19, 22?] in the last simulation steps. The fast convergence of the DoS in the initial simulation steps to the neighborhood of the exact DoS is still not understood, although the Wang-Landau algorithm is widely used.

Calculation with modified WL-1/t algorithm

We use the Wang-Landau algorithm [11] with 1/t-modification [20] and compute convergence criterion δ [19] during simulation.

We simplify the protocol by fixing an interval of M between checks of the Bellardineli and Pereyra $H(E)! = 0$ criterion with $M = 10^5$, which avoids multiple and unnecessary computations. We also use this point to compute the largest eigenvalue of $T(E_k, E_m)$ using the procedure `dgeev()` from the Intel oneAPI Math Kernel Library LAPACK [37]. The random number generator `mt19937.c` [38, 39] was used to generate a new configuration state and decide whether to accept the new configuration.

Equilibrium microcanonical annealing algorithm

A promising framework for simulating equilibrium systems in a microcanonical ensemble using annealing in an energy ceiling has been proposed by Rose and Machta and successfully applied to the first-order thermal transition in a 20-component two-dimensional Potts model with demonstration of topological transitions in the phase coexistence region [23].

Rose-Machta ceiling procedure

Here, we briefly introduce the Rose and Machta ceiling population algorithm presented in Section II of the paper [23].

In Rose and Machta's approach to simulating equilibrium systems in a microcannonical ensemble, there is no relaxation on temperature decrease; instead, the independent variable of the algorithm is energy. The MCMC procedure consists of a single spin-flip. Moves occur in configuration space, changing a randomly chosen spin, and the transition probability from an α state with energy E_α to a ω state with energy E_ω is defined as follows

$$P_{ceiling}(\alpha \to \omega) = \begin{cases} 1 & \text{if } E_\omega \leq E_c \\ 0 & \text{if } E_\omega > E_c \end{cases}, \qquad (12)$$

where E_c is the ceiling energy value, the *cooling* energy value. An elementary MCMC step consists with N updating of randomly chosen spins, where N is the number of spins in the system. The number of elementary MCMC steps $n_s(E)$ is a parameter of the algorithm.

The method is in a sense a mixture of three algorithms: the simulated annealing algorithm [13], the Wang-Landau algorithm [11, 12], and the population annealing algorithm [14, 15].

Floor microcanonical procedure

A practical modification of the Rose-Machta algorithm has been augmented [24] the floor energy in addition to the ceiling energy, allowing the DoS to be estimated for all energy levels of the system. Here we basically repeat the relevant part of the paper [24]. The idea is to generate random configurations of the system that are most likely to correspond to energy levels in the neighborhood of the DoS maximum, which is a convex function. By applying the ceiling energy algorithm, the left wing of the DoS can be estimated as the ceiling algorithm guides the replicas to the ground state. An extension of the ceiling algorithm, the floor algorithm [24], instead directs replicas to a higher energy.

In the floor algorithm, the probability of transition from the α state with energy E_α to the ω state with energy E_ω is defined as follows

$$P_{floor}(\alpha \to \omega) = \begin{cases} 1 & \text{if } E_\omega \geq E_f \\ 0 & \text{if } E_\omega < E_f \end{cases}, \qquad (13)$$

where E_f is the floor energy value.

Microcanonical population annealing algorithm

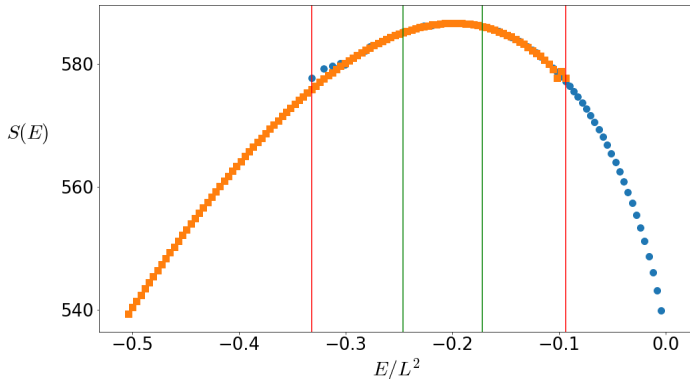Combining the ceiling and floor procedures we get the following algorithm.

FIG. 2: Illustration of the entropy estimation stitching procedure for the 10-component Potts model on a square lattice with linear size L=16. The orange squares correspond to the ceiling process and the blue circles to the floor process. The meaning of the vertical lines discussed in the text.

First we need to prepare a *pool* of replicas, for which we generate R replicas of the system. The random distribution of spins in each replica will case the replica energies to have a value close to the maximum of the energy probability distribution $g(E)$, called the density of states (DOS). The first value of the ceiling E_c will be the highest energy in the replica pool, and the first value of the floor E_f will be the lowest energy in the replica pool.

In general, we do not know the energy spectrum in advance, and the next energy ceiling $E_c(i+1)$ or floor value $E_f(i+1)$ is chosen as the closest to the current energy from the energies of the current replica pool. Thus, we compute the energy spectrum of the system dynamically.

The elementary step i of algorithm is as follows [24]

1. Select the new ceiling (floor) value. This could be the energy level down $E_c(i)$ from the current ceiling procedure, or the energy level up $E_f(i)$ from the current floor procedure.

2. Perform the $n_s(E_c(i))$ or $n_s(E_f(i))$ MCMC check, thereby creating new configurations R that represent a pool of configurations.

3. Count the number of replica in the pool R' with energy $E_c(i)$ or $E_f(i)$, and calculate the culling fraction $\epsilon(E_c(i))=R'/R$ or $\epsilon(E_f(i))=R'/R$. Filter these R' configurations from the pool of configurations.

4. Randomly select with repetitions new replicas R from the filtered pool of configurations.

The process was terminated if the condition $R'=R$ was satisfied, which means that the annealing reached the ground state in the ceiling protocol or the most symmetric energy level in the floor protocol.

Stitching entropy parts together

To estimate the extensive part of entropy [23], the culling fractions for the ceiling and floor are used

$$S^c(E) = \ln(\epsilon(E)) + \sum_{E'>E} \ln(1 - \epsilon(E')), \quad (14)$$

$$S^f(E) = \ln(\epsilon(E)) + \sum_{E'<E} \ln(1 - \epsilon(E')). \quad (15)$$

Entropy allows us to add arbitrary constants, which we denote as S_0^c and S_0^f, the entropy constants for the ceiling and floor, respectively.

As can be observed on a model example of figure 2, both cooling and heating run cover only one wing of the whole spectrum. The intersection, covered by both runs, is placed near the maximum of entropy, where random replica normally reside. We obtain entropy in the full energy range by stitching cooling and heating wings in the intersecting region.

Stitching is somewhat arbitrary procedure and not sensitive to the exact choice of it. We conduct it as follows:

1. Select the intersection area bounded by the red lines in Fig. 2 from the leftmost point of the "heating" wing to the rightmost point of the "cooling" wing.

2. The ends of the cooling and heating wings are somewhat scattered, so we cut off the outer parts of the region, leaving the region bounded by the green lines labeled E_{left} and E_{right}. In this paper, we use a width between green lines three times smaller than between red lines, although this ratio can be changed depending on the specific task.

3. Calculate the average $\Delta S = avg(S^c(E) - S^h(E))$ over all energies in last "green" area. This allows us to write cross-linked $S(E)$ in the form

$$S(E) = S_0 + \begin{cases} S^c(E) & \text{if } E < E_{left} \\ S^h(E) + \Delta S & \text{if } E > E_{right} \\ (S^c(E) + S^h(E) + \Delta S)/2 & \text{else} \end{cases}$$
$$(16)$$

4. The last free constant S_0, if required, can be established by counting the number of all states in the system, which in the case of the q-state Potts model is q^{L^2}, and S_0 should be chosen from the relation

$$\sum_E e^{S(E)} = q^{L^2}. \quad (17)$$

Estimation of the thermodynamic observables

An estimate of the partition function is given as a function of temperature T measured in the energy units

$$Z(T) = \sum_E e^{-E/T + S(E)}. \quad (18)$$

The estimates of the average internal energy $\langle E(T)\rangle$, specific heat $\langle C(T)\rangle$, Binder cimulant [40], and probability of energy distribution of $P(E;T)$ at temperature T are calculated using the following expressions

$$\langle E(T)\rangle = \frac{\sum_E E\, e^{-E/T+S(E)}}{Z(T)}, \quad (19)$$

$$\langle E^2(T)\rangle = \frac{\sum_E E^2\, e^{-E/T+S(E)}}{Z(T)}, \quad (20)$$

$$C(T) = \frac{\langle E^2(T)\rangle - \langle E(T)\rangle^2}{T^2}, \quad (21)$$

$$V(T) = 1 - \frac{\langle E^4(T)\rangle}{3\langle E^2(T)\rangle^2} \quad (22)$$

$$P(E;T) = \frac{e^{-E/T+S(E)}}{Z(T)}. \quad (23)$$

Calculation with MCPA algorithm

The implementation of the MCPA algorithm is based on a modification [23] of the accelerated population annealing algorithm for GPU [41], presented in [24]. The simulations were performed on an NVIDIA V100 GPU with a typical replica number $R=2^{17}=131072$. Spins are represented by a single number of type C 'char'.

We use the cuRAND [42] package with the Philox random number generator from the CUDA SL package, which allows us to obtain independent sequences of pseudo-random numbers. The largest linear lattice size in our study is $L=70$, and about $2 \cdot n_s \cdot L^2 \approx 10^5$ random numbers per algorithm step are used to simulate the ceiling/floor in one replica. The total number of steps is equal to the number of energy levels, which is $2L^2-3 \approx 10^4$. The total number of random numbers per run of one replica is about 2^{30}, which is less than the length of Philox stream 2^{64}.

RESULTS COMPARISON

In this section, we directly compare the simulation results of the Potts model with 10 and 20 components using the WL-1/t algorithm [11, 20] with control of convergence [19] and MCPA algorithm [23, 24].

We will present a comparison of the energy density distribution $P(E)$, estimates of the energies in the ordered and disordered states, estimates of the critical temperature, estimates of the maximum of the specific heat, estimates of the Binder cumulant minimum, and the ratio of $P(E)$ peaks at and near the critical temperature.

Specific heat and Binder cumulant scaling

Finite-size analysis of the specific heat and Binder cumulant was carried out in [43] and extended in [44], and we use these analytical conclusions in the following sections. We estimate critical temperature from the position of the specific heat maximum and the Binder cumulant minimum, the dependence of the magnitude of the specific heat capacity maximum on the lattice size, and the magnitude of the Binder cumulant minimum. Other quantities, such as the value of the coefficients in the L^{-2} corrections to the above estimates, can also be evaluated, but such a detailed analysis is beyond the scope of this paper. A corresponding analysis was presented in [44] for the cases of the 8- and 10-component Potts model.

Specific heat with q=10

Figure 3 shows the specific heat in the critical region of the 10-component Potts model calculated with the WL and MCPA methods and compares it with the analytical estimate given in [44], expr. (3.18). The comparison is very good, as already noted for the 10-component Potts model in the Figure 10 of the paper [44].
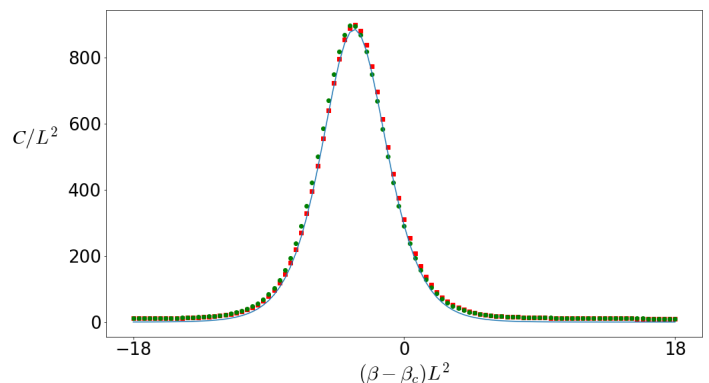

FIG. 3: Specific heat as a function of inverse temperature β in the critical region of the 10-component Potts model near the critical value β_c. Red squares are WL simulations, green circles are MCPA simulations, and the solid line is the analytical expression (3.18) of paper [44]. The linear size of the lattice is $L=60$.

The maximum value of the specific heat C_{max} of the Potts model in the case of a first-order phase transition depends on the volume of the system as [43, 44]

$$C_{max}/L^2 \propto \frac{(e_d - e_o)^2}{4T_c^2} = \alpha_{exact}. \quad (24)$$

Fitting the data in the fourth and fifth columns of Table I yields a scaling estimate of the C_{max} with slope $\alpha \approx 0.2473(3)$ for WL and for MCPA with slope

TABLE I: Estimation of the maximum C_{max} of specific heat and its position T_{Cmax} for $q = 10$.

L	$T_{C_{max}}$		C_{max}	
L	WL	MCPA	WL	MCPA
16	0.70696	0.70696	74.0	73.7
30	0.70299	0.70299	233.3	233.5
40	0.70225	0.70225	406.3	404.7
50	0.70185	0.70190	628.9	628.6
60	0.70166	0.70166	900.6	900.4
∞	0.70122(2)	0.70125(3)		
exact	0.70123...			

TABLE II: Estimation of the maximum C_{max} of specific heat and its position T_{Cmax} for $q = 20$.

L	$T_{C_{max}}$		C_{max}	
L	WL	MCPA	WL	MCPA
30	0.58934	0.58926	930.1	931.4
40	0.58891	0.58886	1651.6	1651.7
50	0.58869	0.58866	2581.9	2577.9
60	0.58864	0.58852	3712.1	3712.6
70	0.58855	0.58845	5049.8	5052.9
∞	0.58837(2)	0.58829(2)		
exact	0.5883498...			

$\alpha \approx 0.2471(4)$, taking into account the correction terms in the fit, Expr. (A23) from the paper [44]. Our estimates agree well with the numerical estimate from earlier paper [43], $\alpha \approx 0.250$, and more closer to the exact value of $\alpha_{exact} = 0.246355...$.

The critical temperature T_c can be estimated from the position $T_{C_{max}}$ of the the maximum of the specific heat, and fitting the data in the second and third columns of the table I gives the estimates shown in the penultimate row of the table. The last row shows the exact known temperature. The estimates match the exact value to within four digits. Note that the estimates for the data obtained by the WL and MCPA methods are equally good.

Specific heat with q=20

Figure 4 shows the specific heat in the critical region of the 20-component Potts model calculated with the WL and MCPA methods and compares it with the analytical estimate given in [44], expr. (4.5).
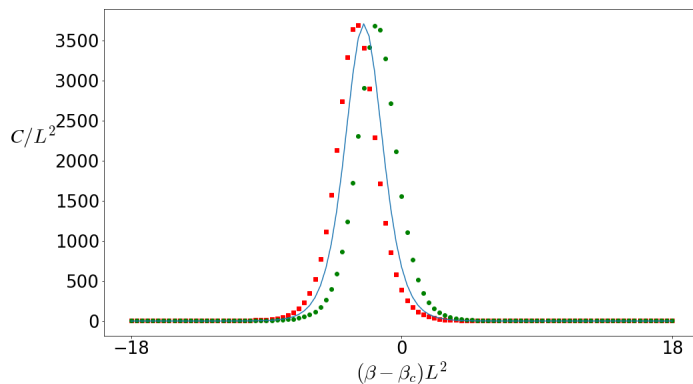
TABLE III: Comparison of B_{min} and β_{B_m} for $q = 10$.

L	$\beta_{B_{min}}$, WL	$\beta_{B_{min}}$, MCPA	B_{min}, WL	B_{min}, MCPA
16	1.4073	1.4072	0.525	0.525
30	1.4206	1.4207	0.548	0.547
40	1.4230	1.4230	0.552	0.553
50	1.4241	1.4241	0.554	0.554
60	1.4247	1.4247	0.555	0.555
∞	1.42614(3)	1.42601(4)	0.5573(4)	0.5573(5)
exact	1.42606...		0.55889...	

table II yeilds a scaling estimate of C_{max} with a slope of $\alpha \approx 1.0306(6)$ for WL and for MCPA with a slope of $\alpha \approx 1.0299(4)$. Our estimates agree well with the analytical prediction [43, 44] $\alpha = 1.02987966...$ calculated from Expr. (24) using exactly known values of the critical temperature and energies of the ordered and disordered phases for the 20 components Potts model, expressions (2-4).

Binder cumulant with q=10

The inverse temperature value $\beta_{B_{min}}$ at the minimum of the Binder cumulant can be used to estimate critical value, which is presented in the second and third columns of Table III for the WL and MCPA methods. The magnitude of the minimum value B_{min} was calculated analytically in [44]

FIG. 4: Same as in Fig. 3, but for the Potts model with 20-components.

Fitting the data in the fourth and fifth columns of

$$B_{min} = \frac{2}{3} - \frac{(e_o/e_d - e_d/e_o)^2}{12} + O(L^{-d}) \qquad (25)$$

and are given in the last row of table III along with the exact value of β_c. The estimates of $\beta_{B_{min}}$ and B_{min} given in the last to next row and obtained with WL and MCPA methods agree well with the exact values.

TABLE IV: Comparison of B_{min} and β_{B_m} for $q = 20$.

L	$\beta_{B_{min}}$, WL	$\beta_{B_{min}}$, MCPA	B_{min}, WL	B_{min}, MCPA
30	1.694792	1.695030	0.1047	0.1032
40	1.696914	1.697078	0.1100	0.1112
50	1.697963	1.698049	0.1129	0.1147
60	1.698332	1.698675	0.1154	0.1166
70	1.698727	1.699004	0.1170	0.1175
∞	1.6995(2)	1.6999(2)	0.1189(7)	0.1209(1)
exact	1.699669...		0.1197...	

Binder cumulant with q=20

The estimates of the inverse temperature value $\beta_{B_{min}}$ and Binder's cumulant minimum from the data in Table IV agree well with the analytically known β_c and B_{min}, Expr. (2) and Expr. (25) for the 20-components Potts model.

Probability distribution of energy

In [44] it is emphasized that in a finite system in the vicinity of a phase transition in the temperature range of order one, $L^d(T-T_C)/T \approx O(1)$, all states contribute significantly to the energy distribution. Therefore, the contribution of q states to the ordered phase will give a peak in the probability distribution of energy $P(E)$ about q times larger than the disordered state (the state with maximal possible symmetry!). Figure 5 shows the probability distribution of the energy $P(E)$ for linear lattice sizes $L = 30$ and 60 for the 10-component Potts model, estimated by the WL and MCPA methods. The ratio of the amplitudes of the peaks $r = P(E)_o^{max}/P(E)_d^{max}$ corresponding to the ordered and disordered phases is given in Table V. In addition, we give estimates of the critical ratio r_c [23]

$$r_c = \frac{\sum_{E<E_c} P(E)}{\sum_{E \geq E_c} P(E)} \qquad (26)$$

taken with the energy distribution at critical temperature T_c and with the sums divided by the critical energy $E_c = (E_o + E_d)/2$.

The results in Table V show that both computational methods lead to reasonable estimates of this ratio. As can be seen in Figure 6 and Table VI, this is not as good for large lattice sizes in the case of the 20-component Potts model. This is because at larger lattice sizes the distributions become narrower, and the results are exponentially sensitive [46] to small variations in the computations. The same effect can be seen in the figure 4 for specific heat in the case of the 20-component model, where the maximum of specific heat is very close for the

WL and MCPA modeling case and the analytical approximation, and this is noticeable despite the fact that the difference is only in the fourth digit (see Table II) and simply because of the tiny scale of the peak.
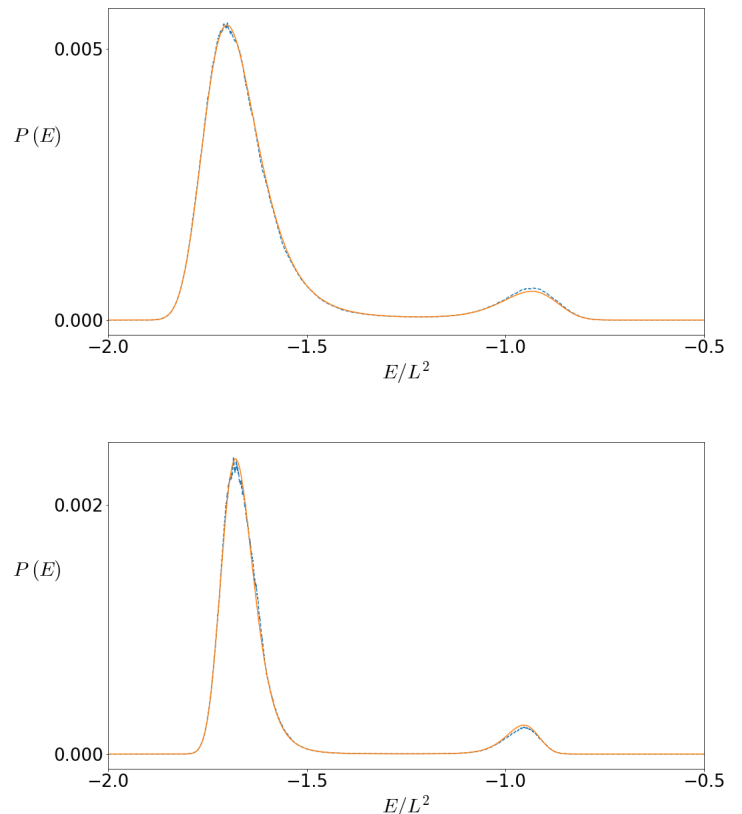

FIG. 5: Energy probability distribution $P(E)$ for 10-component Potts model for two lattice size - top: $L = 30$ and bottom: $L = 60$. The solid orange line is estimation by the WL method, the dashed blue line - by MCPA method.

TABLE V: Comparison for peak ratio and r_c at exact critical temperature, $q = 10$.

L	Ratio, WL	Ratio, MCPA	r_c, WL	r_c, MCPA
16	10.5131	10.6(3)	7.9709	8.1(2)
30	10.2975	10.5(7)	9.2050	9.3(6)
40	10.2091	10.7(7)	9.5976	10.0(6)
50	10.0828	10.7(9)	9.6729	10.2(8)
60	10.2101	10(1)	9.9473	9.5(9)

Estimation of energies of ordered and disordered states

An estimate of the latent heat can be obtained from the positions $e_0(L)$, $e_d(L)$ of the maximum $P(e)$ calculated at the exact critical temperature. Table VII and VIII
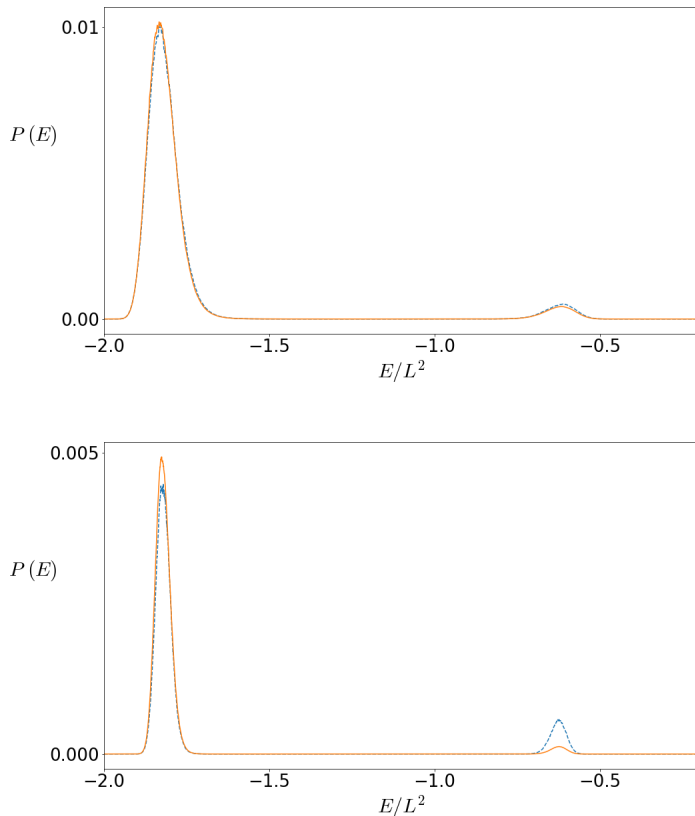
FIG. 6: Similar to Fig. 5 for the 20-component Potts model.

TABLE VII: Estimation of e_0 and e_d from the position of the maximum $P(e)$ calculated at the exact critical temperature, $q = 20$. The penultimate entry in the table is a fit to data.

L	E_0, WL	E_0, MCPA	E_d, WL	E_d, MCPA
6	-1.74219	-1.74609	-0.89844	-0.89062
30	-1.70000	-1.70000	-0.93444	-0.92667
40	-1.69063	-1.69938	-0.94375	-0.94812
50	-1.68360	-1.68560	-0.94960	-0.93920
60	-1.67944	-1.68472	-0.95444	-0.95028
∞	-1.657(1)	-1.661(4)	-0.9743(5)	-0.972(6)
exact	-1.664		-0.968	

TABLE VIII: As in the table VII for the 20-component Potts model.

L	e_0, WL	e_0, MCPA	e_d, WL	e_d, MCPA
30	-1.83333	-1.83333	-0.61556	-0.61222
40	-1.82750	-1.82937	-0.62125	-0.62375
50	-1.82440	-1.82680	-0.62200	-0.61840
60	-1.82556	-1.82056	-0.62389	-0.62722
70	-1.82245	-1.81939	-0.62408	-0.62592
∞	-1.820(1)	-1.8173)	-0.6272(4)	-0.631(3)
exact	-1.82068		-0.62653	

TABLE VI: Comparison for peak ratio and r_c at exact critical temperature, $q = 20$.

L	Ratio, WL	Ratio, MCPA	r_c, WL	r_c, MCPA
30	23.7	18(1)	21.8	17(1)
40	24.3	20(3)	22.4	18(3)
50	20.5	20(5)	19.0	19(5)
60	39.9	12(3)	36.5	11(3)
70	31.3	60(30)	29.0	55(30)

summarize the positions of the peaks. We find that the data fit well with ansatz of the form $e_0 = e_0(L) + b/L$ with $L \to \infty$ and $e_d = e_d(L) + b/L$ with $L \to \infty$. The fitting results give fairly good estimates of e_0 and e_d compared to the exactly known results given in the last rows of the tables.

DISCUSSION

In this paper, two methods for direct estimation of the density of states (DOS) in the spin model of statistical mechanics with discrete spectrum are comparatively analyzed. First, these are computational results for the modified Wang-Landau [20] algorithm, a 1/t-WL algorithm, with accuracy control [19]. Second, these are computational results for the microcanonical population annealing algorithm [23, 24], the MCPA algorithm. MCPA is a new algorithm, and not much is known about its properties. Therefore, a direct comparison of the two algorithms is of some interest.

We stop simulation with 1/t-WL after the parameter f reaches the value f_{fin}=1.00000001, as proposed in the original version of WL algorithm [11]. We control the convergence of DoS by computing the largest eigenvalue of the transfer matrix in the energy spectrum [19] (see Figures 1 and 2). The number of spin update operations ranges from $5 \cdot 10^{10}$ to $72 \cdot 10^{10}$ for lattice sizes from L=16 to L=60.

The MCPA simulation is completed using a ceiling/floor procedure, reaching the lowest/highest energy level. The number of operations in this case is not random. The number of spin updates in each replica is equal to the product of the number of spins L^2, the number of MCMC iterations n_s, and the number of energy levels N_E=$2L^2-3$ for the 10- and 20-component models. For a typical number of iterations n_s=10, the number of spin flips per replica varies from $16 \cdot 10^6$ to $480 \cdot 10^6$, and the linear size of the system varies from L=30 to L=70. The total number of spin updates must be multiplied by the number of replicas R=$2^{17} \approx 1.3 \cdot 10^5$. Thus, the total number of spin updates is comparable to an order-of-magnitude precision.

The estimates of thermodynamic quantities computed

with the WL and MCPA algorithms have the same accuracy, as shown in the paper, and the advantage of MCPA is that it requires fewer operations per thread than the WL algorithm. Of course, the threads in both cases are different: WL uses CPU threads, while MCPA uses GPU threads. Therefore, it is nor easy way to compare computational effectiveness of algorithms.

In addition, MCPA can be used to estimate the magnetization and its moments, i.e., magnetic susceptibility and magnetic Binder cumulant [40] from a single run and averaging as functions of the replica pool for energy E' (the culling pool), yielding $M(E')$, $\xi(E')$, and $B(E')$. In contrast, to extract the magnetic sector of thermodynamic variables using the WL algorithm, the two-dimensional distribution function $g(E, M)$ [45] must be evaluated for both energy and magnetization, which significantly increases the required computation time. This is one of the advantages of MCPA over WL.

---

[1] I.G. Enting, J. Guttmann, I. Jensen, *Low-temperature series expansions for the spin-1 Ising model*, J. Phys.A **27** (1994) 6987.

[2] P. Butera and M. Pernici, *High-temperature expansions of the higher susceptibilities for the Ising model in general dimension d*, Phys. Rev. E **86** (2012) 011139.

[3] H.W.J. Blóte and M.P. Nightingale, *Critical behaviour of the two-dimensional Potts model with a continuous number of states; A finite size scaling analysis*, Physica A **112A** (1982) 405.

[4] B. Derrida, J. Vannimenus, *Transfer-matrix approach to percolation and phenomenological renormalization*, Journal de Physique **41** (1980) 473.

[5] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press: Cambridge, 2021).

[6] C.J. Geyer, *Practical Markov chain Monte Carlo*, Stat. Sci. **7** (1992) 473.

[7] R.H. Swendsen and J.-S. Wang, *Nonuniversal Critical Dynamics in Monte Carlo Simulations*, Phys. Rev. Lett. **58** (1987) 86.

[8] U. Wolff, *Collective Monte Carlo updating for spin systems*, Phys. Rev. Lett. **62** (1989) 361.

[9] W. Janke, *Multicanonical Monte Carlo simulations*, Physica A **254** (1998) 164.

[10] P. Kar, W. Nadler, and U.H.E. Hansmann, *Microcanonical replica exchange molecular dynamics simulation of proteins*, Phys. Rev. E **80** (2009) 056703.

[11] F. Wang and D.P. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states*, Phys. Rev. Lett. **86** (2001) 2050.

[12] F. Wang and D.P. Landau, *Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram*, Phys. Rev E **64** (2001) 056101.

[13] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Optimization by simulated annealing*, Science **220** (1983) 671.

[14] J. Machta, *Population annealing with weighted averages: A Monte Carlo method for rough free-energy landscapes*, Phys. Rev. E **82** (2010) 026704.

[15] M. Weigel, L.Yu. Barash, L.N. Shchur and W. Janke, *Understanding population annealing Monte Carlo simulations*, Phys. Rev. E **103** (2021) 053301.

[16] J. Hénin, T. Leliévre, M.R. Shirts, O. Valsson, and L. Delemotte, *Enhanced sampling methods for molecular dynamics simulations*, Living J. Comput. Mol. Sci. **4** (2022) 1583.

[17] S. Wansleben and D.P. Landau, *Monte Carlo investigation of critical dynamics in the three-dimensional Ising model*, Phys. Rev. B **43** (1991) 6006.

[18] X.-J. Li and A.D. Sokal, *Rigorous lower bound on the dynamic critical exponents of the Swendsen-Wang algorithm*, Phys. Rev. Lett. **63** (1989) 827.

[19] L.Y. Barash, M.A. Fadeeva and L.N. Shchur, *Control of accuracy in the Wang-Landau algorithm*, Phys. Rev. E **96** (043307) 6006.

[20] R.E. Belardinelli and V.D. Pereyra, *Fast algorithm to calculate density of states*, Phys. Rev. E **75** (2007) 046701.

[21] F. Liang, *A theory on flat histogram Monte Carlo algorithms*, J. Stat. Phys. **122** (2006) 511.

[22] F. Liang, C. Liu, and R. J. Carroll, *Stochastic approximation in Monte Carlo computation*, J. Am. Stat. Ass. **102** (2007) 305 .

[23] N. Rose and J. Machta, *Equilibrium microcanonical annealing for first-order phase transitions*, Phys. Rev. E **100** (2019) 063304.

[24] V.K. Mozolenko, and L.N Shchur, *Blume-Capel model analysis with a microcanonical population annealing method*, Phys. Rev. E **109** (2024) 045306 .

[25] P. Schierz, J. Zierenberg, and W. Janke, *First-order phase transitions in the real microcanonical ensemble*, Phys. Rev. E **94** (2016) 021301(R).

[26] W. Janke, P. Schierz, and J. Zierenberg, *Transition barrier at a first-order phase transition in the canonical and microcanonical ensemble*, J. Phys.: Conf. Ser. **921** (2018) 012018.

[27] R.B. Potts, *Some generalized order-disorder transformations*, Proc. Cambridge Philos. Soc. **48** (1952) 106.

[28] F.Y. Wu, *The Potts model*, Rev. Mod. Phys. **54** (1982) 235.

[29] R.J. Baxter, *Exactly solved models in statistical mechanics*, (London: Academic Press, 1982).

[30] M.P. Taylor, W. Paul, and K. Binder, *Phase transitions of a single polymer chain: A Wang–Landau simulation study*, J. Chem. Phys. **131** (2009) 114907.

[31] D.T. Seaton, T. Wüst, and D.P. Landau, em Collapse transitions in a flexible homopolymer chain: Application of the Wang-Landau algorithm, Phys. Rev. E **81** (2010) 011802.

[32] N.A. Taklimi, F. Ferrari, M.R. Piatek, and L. Tubiana, *Thermal properties of knotted block copolymer rings with charged monomers subjected to short-range interactions*, Phys. Rev. E **108** (2023) 034503.

[33] H. Bian, X. Shao, W. Cai, and H. Fu, *Understanding the Reversible Binding of a Multichain Protein–Protein Complex through Free-Energy Calculations*, J. Chem. Phys. **128** (2024) 3598.

[34] H.K. Lee, Y. Okabe, D.P. Landau, *Convergence and refinement of the Wang-Landau algorithm*, Comp. Phys. Comm. **175** 2006 36.

[35] C. Zhou and R.N. Bhatt, *Understanding and improving the Wang-Landau algorithm*, Phys. Rev. E **72** (2005) 025701(R).

[36] P.D. Beale, *Exact distribution of energies in the two-dimensional Ising model*, Phys. Rev. Lett. **76**(1996) 78.

[37] M. Krainiuk, M. Goli, and V.R. Pascuzzi, oneAPI Open-Source Math Library Interface, 2021 International Workshop on Performance, Portability and Productivity in HPC (P3HPC), p. 22.

[38] L.Yu. Barash and L.N. Shchur, *PRAND: GPU accelerated parallel random number generation library: Using most reliable algorithms and applying parallelism of modern GPUs and CPUs*, Comp. Phys. Comm. **185** (2014) 1343.

[39] M.S. Guskova, L.Yu. Barash and L.N. Shchur, *RN-GAVXLIB: Program library for random number generation, AVX realization*, Comp. Phys. Comm. **200** (2016) 402.

[40] K. Binder, *Finite size scaling analysis of Ising model block distribution functions*, Z. Phys. B **43** (1981) 119.

[41] L. Y. Barash, M. Weigel, M. Borovský, W. Janke, and L. N. Shchur, *GPU accelerated population annealing algorithm*, Comput. Phys. Commun. **220** (2017) 341.

[42] *cuRAND*, CUDA Toolkit v12.3.0, Developer zone. https://docs.nvidia.com/cuda/curand/device-api-overview.html

[43] M.S.S. Challa, D.P. Landau, and K. Binder, *Finite-size effects at temperature-driven first-order transition*, Phys. Rev. B **34** (1986) 1841.

[44] J. Lee and J.M. Kosterlitz, *Finite-size scaling and Monte Carlo simulations of first-order phase transitions*, Phys. Rev. B **43** (1991) 3265.

[45] C. Zhou,T. C. Schulthess, S. Torbrügge, and D. P. Landau, *Wang-Landau algorithm for continuous models and joint density of states*, Phys. Rev. Lett. **96** (2006)120201.

[46] In some cases, the least significant digits can be important. According to the memories of colleagues, Lev Landau used to say that "money remains in an exponential function", so counting it to the last coin is exponentially important.

[47] The references cited in this paragraph reflect the authors' subjective view on the topic of modeling classical spin models and do not pretend to cover the entire wide range of the many good and informative papers.